International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 19(4): 223–232 (2010)
Published online 18 August 2010 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/mpr.323

Identifying HAM-A cutoffs for mild, moderate, and severe generalized anxiety disorder

LOUIS S. MATZA,[1] ROBERT MORLOCK,[2] CHRIS SEXTON,[1] KAREN MALLEY[3] & DOUGLAS FELTNER[4]

1 Center for Health Outcomes Research, United BioSource Corporation, Bethesda, MD, USA
2 i3 Innovus, Eden Prairie, MN, formerly with Pfizer Inc., Ann Arbor, MI, USA
3 Malley Research Programming, Inc., Rockville, MD, USA
4 Pfizer, Inc., Groton, CT, USA

Key words
HAM-A, anxiety, cutoffs, severity ranges, generalized anxiety disorder

Correspondence
Louis S. Matza, United BioSource Corporation, 7101 Wisconsin Avenue, Suite 600, Bethesda, MD 20814, USA.
Telephone (+1) 301-664-7263
Fax (+1) 301-654-9864
Email: louis.matza@unitedbiosource.com

Received 19 November 2008; revised 3 June 2009; accepted 22 July 2009

Abstract
The aim of the current study was to identify and evaluate cutoffs for mild, moderate, and severe ranges of Hamilton Anxiety Rating Scale (HAM-A) scores. Data were from a four-week randomized trial of treatment for generalized anxiety disorder. Measures included the HAM-A, SF-36, Hospital Anxiety and Depression Scale (HADS), and Clinical Global Impressions of Severity (CGI-S) scale. HAM-A cutoffs were identified based on literature review, expert panel input, and MANOVA models. The optimal cutoff set was evaluated based on association with clinician CGI-S ratings. The sample included 144 patients (56.3% female; 73.6% white; mean age = 35.7 years; mean baseline HAM-A score = 23.7). The optimal HAM-A score ranges were: mild anxiety = 8–14; moderate = 15–23; severe $\geq$ 24 (scores $\leq$ 7 were considered to represent no/minimal anxiety). Analysis of variance (ANOVA) models found statistically significant differences among these groups in the SF-36 and HADS. The HAM-A severity ranges closely corresponded to clinicians' CGI-S ratings. The study represents the first step towards developing severity ranges for the HAM-A. These cutoffs should be used with caution and validated in larger samples. If the proposed cutoffs are accepted for general use, they could make results more meaningful and interpretable for researchers, clinicians, and patients.

Introduction

The Hamilton Anxiety Rating Scale (HAM-A) was originally developed over 40 years ago as a clinician-rated instrument for quantifying anxiety symptoms (Hamilton, 1959). The 14-item version remains the most commonly used outcome measure in clinical trials of treatments for anxiety disorders (Allgulander *et al.*, 2004; Davidson *et al.*, 2005; Hamilton, 1969; Lenox-Smith and Reynolds, 2003; Lenze *et al.*, 2005; Mathew *et al.*, 2005; Naukkarinen *et al.*, 2005; Pollack *et al.*, 2002; Rickels *et al.*, 2005). The HAM-A is also used to assess anxiety symptoms in studies of treatments for other psychiatric and medical conditions, most commonly major depression (Calabrese *et al.*, 2005; Debattista *et al.*, 2005; Gulseren *et al.*, 2006; Musselman *et al.*, 2006; Shelton *et al.*, 2001; Smeraldi, 1998; Wellington and Perry, 2001).

Some researchers have observed limitations of the HAM-A. For example, the instrument may not sufficiently

discriminate between symptoms of depression and anxiety (Maier *et al.*, 1988; Riskind *et al.*, 1987). Furthermore, HAM-A scores may be influenced by somatic medication side effects because the instrument includes several items assessing somatic symptoms (Maier *et al.*, 1988). Despite these limitations, the HAM-A has been able to detect treatment effects in numerous trials, and it has demonstrated adequate reliability, validity, and sensitivity to change (Bech *et al.*, 1984; Clark and Donovan, 1994; Kellner *et al.*, 1968; Maier *et al.*, 1988; Shear *et al.*, 2001). The large body of published HAM-A data is another strength of the instrument because findings can be compared across trials, and data from multiple trials can be pooled for meta-analyses (Wan *et al.*, 2006). Given the quantity of existing HAM-A data and the likelihood that the measure will continue to be used, further research is needed to aid in the interpretation of HAM-A findings.

Thus, the aim of the current study was to propose and evaluate cutoffs for severity ranges of the HAM-A total score. Score cutoffs for mild, moderate, and severe ranges are useful guidelines for interpreting data in individual studies. Furthermore, when cutoffs become widely accepted by researchers, they can facilitate communication and standardize interpretation of remission, study entry criteria, diagnostic criteria, and disease severity (Aben *et al.*, 2002; Bagby *et al.*, 2004; Moller, 2001). Careful identification of cutoffs is critical, as the resulting score ranges have a direct impact on interpretation of treatment outcome (Zimmerman *et al.*, 2004).

In previous studies, HAM-A scores ≤ 7 have generally been thought to indicate remission of anxiety, whereas minimum scores of roughly 17 to 21 are usually required for inclusion in a clinical trial of treatment for generalized anxiety disorder (GAD) (Allgulander *et al.*, 2004; Belzer and Schneier, 2006; Lenze *et al.*, 2005; Llorca *et al.*, 2002; Mathew *et al.*, 2005; Montgomery *et al.*, 2002; Rickels *et al.*, 2003; Sheehan, 2001; Wan *et al.*, 2006). However, standardized score ranges for interpretation of anxiety severity have not previously been proposed. In the current study, severity score ranges were suggested based on clinical expert input and empirical analysis. Then, patients were categorized into severity groups, and the groups were examined to assess whether the cutoffs resulted in distinct groups that correspond to clinicians' perceptions of the patients.

Materials and methods

Data source

This study used data from a four-week, double-blind, randomized, multicenter, placebo-controlled clinical trial of treatment for GAD. Data were pooled from patients in the three treatment groups: lorazepam (1.5 mg t.i.d.), paroxetine (20 mg qd), and placebo, all administered orally. Patients were required to be 18 to 65 years old, with a primary diagnosis of GAD [Diagnostic and Statistical Manual of Mental Disorders – 4th edition (DSM-IV), 300.02] as determined by the Mini-International Neuropsychiatric Interview, version 5.0.0 (Sheehan *et al.*, 1998). In addition, patients were required to have a HAM-A score of ≥ 20 at screening (one week prior to baseline) and baseline; Covi Anxiety Scale total score > 9; and Raskin Depression Scale total score < 7 at screening to ensure predominance of anxiety symptoms over depression symptoms. Patients were excluded from the study if they had the following current DSM-IV Axis I diagnoses: major depressive disorder, panic disorder, acute stress disorder, obsessive-compulsive disorder, dissociative disorder, post-traumatic stress disorder, anorexia, social phobia, bulimia, caffeine-induced anxiety disorder, or alcohol/substance abuse/dependence. Patients were also excluded if they had past or current diagnoses of schizophrenia; psychotic disorder; delirium, dementia, or other clinically significant cognitive disorders; bipolar disorder; or schizoaffective disorder.

The current analyses used data from one week after baseline (i.e. week 1) and one week following the end of the trial (i.e. week 5). These data points were chosen because, at these two time points, patients were most evenly distributed across the range of HAM-A severity levels. Data from baseline (i.e. week 0) and endpoint (i.e. week 4) were not appropriate for the current analysis because patients were not distributed across the full range of HAM-A severity levels (i.e. mild, moderate, and severe). At baseline, there were no patients in the mild range, while at the endpoint, most patients had improved, resulting in an insufficient number of patients in the severe range.

A total of 167 patients were enrolled in the study. To qualify for the current analyses, patients were required to have SF-36 and HAM-A data at week 1. A total of 144 patients met these criteria.

Measures

The Hamilton Anxiety Rating Scale (HAM-A)

The HAM-A is a clinician-rated instrument for quantifying anxiety symptoms (Hamilton, 1959, 1969). The commonly used 14-item version was administered in the current study, with higher scores indicating greater anxiety symptom severity. The HAM-A is described in detail earlier.

Int. J. Methods Psychiatr. Res. 19(4): 223–232 (2010). DOI: 10.1002/mpr

SF-36 Health Survey

This 36-item survey was created to collect health status information across a variety of medical conditions (Ware and Sherbourne, 1992). The SF-36 consists of eight subscales (physical functioning, role-physical, bodily pain, general health, vitality, social functioning, role-emotional, mental health) and two overall summary scores: the physical component summary (PCS) score and the mental component summary (MCS) score (McHorney *et al.*, 1994; Ware *et al.*, 1993).

Clinical Global Impressions of Severity (CGI-S) scale

Clinicians rated GAD symptom severity using this single-item global scale with seven response options. The response options are listed in Table 4.

Hospital Anxiety and Depression Scale (HADS)

The Hospital Anxiety and Depression Scale (HADS) is a brief, self-administered questionnaire, consisting of two subscales, one measuring anxiety (HADS-A), the other measuring depression (HADS-D). Items assess symptoms such as anxious mood, restlessness, anxious thoughts, and panic attacks. Each item is rated on a scale ranging from zero (no presence of anxiety) to three (severe feelings of anxiety). Higher scores indicate more severe anxiety or depression (Zigmond and Snaith, 1983).

Statistical procedures: identifying cutoffs for HAM-A severity ranges

Prior to conducting statistical analyses, the lower HAM-A cutoff was determined based on a review of the literature. The cutoff of HAM-A ≤ 7 is the most commonly used threshold for remission of anxiety in clinical trials (Allgulander *et al.*, 2004; Belzer and Schneier, 2006; Lenze *et al.*, 2005; Llorca *et al.*, 2002; Mathew *et al.*, 2005; Montgomery *et al.*, 2002; Rickels *et al.*, 2003; Sheehan, 2001; Wan *et al.*, 2006). Thus, patients with a HAM-A score ≤ 7 were considered to have no or minimal anxiety, and subsequent empirical analysis designed to identify mild, moderate, and severe score ranges of the HAM-A focused on patients with HAM-A scores of at least 8.

Several previous studies have used a statistical approach to identify cutoffs for symptom severity ranges of other measures for use in patients with conditions other than anxiety (Jensen *et al.*, 2001; Paul *et al.*, 2005; Serlin *et al.*, 1995; Zelman *et al.*, 2003). As in these previous studies, the current study used the following three-step approach. First, all potential combinations of cutoffs for ranges of mild, moderate, and severe symptoms were determined. In the current study, 12 potential sets of cutoffs were examined.

Second, using week 1 data, a series of multivariate analysis of variance (MANOVA) models was conducted, one model for each set of proposed cutoffs. In each model, the independent variable is the proposed three-level (mild, moderate, severe) anxiety severity group variable, and the multiple dependent variables represent symptoms and their functional impact. For example, in previous studies designed to identify severity ranges on a pain measure, the dependent variables quantified the symptoms and impact of pain (Paul *et al.*, 2005; Serlin *et al.*, 1995; Zelman *et al.*, 2003, 2005). For the current study, nine dependent variables were included in each MANOVA: the HADS-A (which represented anxiety symptoms) and the eight domain scores of the SF-36 (which represented functional impact). Essentially, these MANOVAs evaluate the ability of each set of cutoffs to distinguish among the three severity groups with regard to the combination of the dependent variables.

Third, after the MANOVAs were run, the 12 sets of cutoffs were compared using three test statistics assessing between-group effects (i.e. the distinctness or degree of difference among the mild, moderate, and severe anxiety groups). The three statistics are Pillai's trace, Wilks' lambda, and Hotelling's trace (Marascuilo and Levin, 1983). The distributions of these criteria are different from each other, but all three can be transformed to the F distribution, which simplifies the comparison among cutoffs sets. As in previous studies using this methodology, larger F values were interpreted as an indication of greater differences among the mild, moderate, and severe groups. Therefore, when comparing among the 12 sets of cutoffs, the set with the largest F statistics was considered to be optimal. Consistency among the three multivariate test criteria reinforces the appropriateness of the cutoff set (Serlin *et al.*, 1995).

Statistical procedures: evaluating cutoffs for HAM-A severity ranges

After the optimal cutoff set was selected based on MANOVA models, additional analyses were conducted to evaluate whether this set of cutoffs categorized patients into groups with meaningful and statistically significant differences from each other. First, analysis of variance (ANOVA) models with Scheffe's *post hoc* pairwise comparisons were conducted to compare the three anxiety groups with respect to outcome variables at week 1 (SF-36, HADS-A) and week 5 (HADS-A).

Second, the chosen set of HAM-A cutoffs was crossed with clinician ratings of GAD symptom severity on the CGI-S in order to examine concordance between the two ways of categorizing patients. This cross-tabulation was performed using data from week 1 and week 5.

Input from expert panel

An expert panel, consisting of six advisors with extensive clinical and research experience in anxiety disorders, was consulted during this study. The number of patients with anxiety disorders that they treated weekly ranged from zero (one advisor was exclusively involved in research and one was retired) to 20, with a mean of nine per week among those involved in clinical work. On average, the advisors had authored or co-authored about 120 peer-reviewed manuscripts focusing on anxiety disorders (with a range of six to 300 published manuscripts).

The six clinicians were consulted during individual telephone interviews at two points in the study. First, they were contacted prior to completion of the analyses. During this initial contact, they were introduced to the study aims and the proposed empirical strategy for identifying severity cutoffs. At this time, the advisors were asked to provide feedback on the methods, the lower cutoff of $\geq$7, and the usefulness of HAM-A severity ranges. There was unanimous support for the lower cutpoint $\leq$7, and all of the advisors agreed that severity ranges with empirically derived cutoffs would aid in the interpretation of HAM-A scores. One of the advisors provided guidance on the dependent variables that would be used in the MANOVAs. The advisors were contacted a second time after the MANOVAs were completed. During this second consultation, the advisors were asked to comment on the sets of cutoffs that were identified.

Results

Sample demographics

A total of 144 GAD patients met criteria for inclusion in the analysis sample. This sample was 56.3% female (n = 81) and 73.6% white (n = 106), with a mean age of 35.7 years (Table 1). The mean HAM-A score at baseline was 23.7. At baseline, over two-thirds (n = 98, 68.1%) of participants had CGI-S score of four, indicating 'moderately ill.' Patients were roughly evenly divided among the three treatment groups, which were pooled for all subsequent analyses. There were no significant differences in demographic or clinical characteristics between participants in the analysis sample (n = 144) and participants who were excluded from the analysis sample because of missing SF-36 or HAM-A data (n = 23).

Table 1 Baseline demographic and clinical characteristics

Characteristic	Analysis sample[a] (N = 144)
Age	
Mean (standard deviation, SD)	35.7 (11.4)
Gender (n, %)	
Male	63 (43.8%)
Female	81 (56.3%)
Race (n, %)	
White	106 (73.6%)
Black	9 (6.3%)
Hispanic	19 (13.2%)
Asian or Pacific Islander	5 (3.5%)
Other	5 (3.5%)
Baseline HAM-A	
Mean (SD)	23.7 (3.5)
Range	20–39
CGI-S score, baseline	
Mildly ill	4 (2.8%)
Moderately ill	98 (68.1%)
Markedly ill	42 (29.2%)
Treatment group (n, %)	
Placebo	53 (36.8%)
Paroxetine	47 (32.6%)
Lorazepam	44 (30.6%)

[a] To be included in the analysis sample, patients had to have SF-36 and HAM-A data at week 1 (i.e. one week following baseline).

Identifying cutoffs for HAM-A severity ranges

Results of the MANOVAs indicated that two sets of cutoffs tended to yield the greatest F values, indicating better discrimination among mild, moderate, and severe groups than the other sets of cutoffs: set 1 (mild 8–14, moderate 15–23, severe $\geq$ 24) and set 2 (mild 8–15, moderate 16–23, Severe $\geq$ 24) (Table 2). Similarly strong values were observed for these sets of cutoffs, with some small differences across the test criteria. While Wilk's Lambda F value was 6.75 for both solutions, set 1 had a slightly higher Hotelling's trace F value (7.77 versus 7.53), and set 2 demonstrated a slightly higher Pillai's trace F value (5.99 versus 5.79). Because these two cutoff solutions had such similar results, members of the expert panel were asked whether they had a preference for one set over the other. Five of the six panel members were available for feedback, all of whom stated that both cutoff sets were reasonable in light of the current empirical support and their previous experience with the HAM-A. Three of the advisors stated a slight

Int. J. Methods Psychiatr. Res. 19(4): 223–232 (2010). DOI: 10.1002/mpr

Table 2 MANOVA of HAM-A cutoffs at week 1, with SF-36 domain scores and HADS-A total score in model

Set number	HAM-A cutoffs for mild, moderate, and severe groups	HAM-A severity group[a]			Wilk's lambda		Pillai's trace		Hotelling's trace	
		Mild, *n*	Moderate, *n*	Severe, *n*	Value	*F*	Value	*F*	Value	*F*
1	8–14, 15–23, ≥24	32	80	25	0.46	6.75	0.58	5.79	1.12	7.77
2	8–15, 16–23, ≥24	37	75	25	0.46	6.75	0.60	5.99	1.08	7.53
3	8–16, 17–23, ≥24	45	67	25	0.48	6.20	0.56	5.43	1.01	7.01
4	8–17, 18–23, ≥24	60	52	25	0.48	6.25	0.56	5.44	1.02	7.08
5	8–14, 15–24, ≥25	32	87	18	0.49	6.06	0.55	5.30	0.98	6.84
6	8–15, 16–24, ≥25	37	82	18	0.49	6.07	0.56	5.49	0.96	6.68
7	8–16, 17–24, ≥25	45	74	18	0.51	5.61	0.52	4.97	0.90	6.27
8	8–17, 18–24, ≥25	60	59	18	0.50	5.77	0.52	5.01	0.94	6.56
9	8–14, 15–25, ≥26	32	92	13	0.51	5.65	0.53	5.05	0.90	6.28
10	8–15, 16–25, ≥26	37	87	13	0.51	5.70	0.54	5.21	0.89	6.20
11	8–16, 17–25, ≥26	45	79	13	0.53	5.29	0.51	4.79	0.83	5.81
12	8–17, 18–25, ≥26	60	64	13	0.51	5.58	0.53	5.02	0.88	6.15

[a] Total $N = 137$. Seven of the 144 patients with HAM-A score ≤ 7 were excluded from this analysis because they did not fit into the mild, moderate, or severe groups.

preference for set 1, and two advisers indicated a slight preference for set 2. Based on the MANOVA results and the advisors' input, subsequent analyses proceeded with evaluating cutoff set 1, although it is unlikely that there are meaningful differences between the two sets. Using cutoff set 1, patients were categorized based on HAM-A severity levels at week 1: 32 participants with mild anxiety, 80 participants with moderate anxiety, and 25 participants with severe anxiety. An additional seven patients with HAM-A scores ≤ 7 were categorized in a no/minimal anxiety group.

Evaluating the HAM-A severity ranges: group comparisons of scores on patient-reported outcome measures

A series of ANOVAs was conducted to compare the mild, moderate, and severe groups at week 1 with respect to scores on the SF-36 and HADS (Table 3). The no/minimal anxiety group was not included in these analyses because there were so few patients in this group. In general, results indicated that the three HAM-A groups were distinct, with statistically significant differences among the three groups in all scales. In most physical scales of the SF-36 (i.e. General Health Perceptions, Physical Functioning, Role – Physical, Physical Component Score), there were significant differences between the mild and moderate groups, as well as between the mild and severe. For all other scales (i.e. psychosocial scales of the SF-36, HADS-A, HADS-D), all three groups were significantly different from each other (all $p < 0.01$). All differences between groups were in the expected direction, with greater HAM-A severity associated with decreased quality of life as measured by the SF-36, as well as increased symptoms as measured by the HADS-A and HADS-D. Results for the HADS-A and HADS-D follow the same pattern at week 5.

Evaluating the HAM-A severity ranges: concordance with clinical global ratings of severity (CGI-S)

Cross tabulations of the correspondence between HAM-A severity groups and clinicians' CGI-S ratings of symptom severity at weeks 1 and 5 are presented in Table 4. Findings indicate close correspondence of the HAM-A severity levels with clinicians' perceptions. At week 1, both patients rated by clinicians as 'not at all ill' were in the no/minimal anxiety HAM-A group. Patients rated by clinicians as 'borderline ill' were evenly divided between the HAM-A no/minimal symptoms group ($n = 4$) and the mild anxiety group ($n = 4$). Participants whom clinicians considered to

Table 3 Analysis of variance comparing mild, moderate, and severe HAM-A groups at week 1 and week 5

Measure	HAM-A severity groups[a]			Overall *F* value	Significant pairwise comparisons[b]
	Mild ($n = 32$), mean (SD)	Moderate ($n = 80$), mean (SD)	Severe ($n = 25$), mean (SD)		
SF-36 scales (week 1)					
General Health Perceptions	76.8 (17.6)	71.2 (16.0)	50.2 (23.4)	17.38***	B***, C***
Physical Functioning	93.8 (10.2)	90.7 (13.7)	76.2 (23.2)	10.93***	B***, C***
Role – Physical	81.8 (20.6)	75.7 (21.7)	49.6 (29.0)	15.91***	B***, C***
Role – Emotional	77.3 (17.2)	60.1 (20.8)	38.3 (22.3)	25.83***	A***, B***, C***
Social Functioning	80.5 (18.8)	58.6 (23.6)	33.0 (20.1)	32.89***	A***, B***, C***
Mental Health Index	69.7 (15.6)	52.9 (17.3)	33.2 (14.6)	34.55***	A***, B***, C***
Pain	83.5 (19.4)	70.2 (22.1)	56.2 (26.4)	10.49***	A*, B***, C*
Vitality	51.2 (20.5)	42.5 (16.1)	24.3 (12.7)	18.87***	A*, B***, C***
Physical Component Score	55.3 (7.2)	55.1 (7.4)	48.9 (10.6)	6.22**	B*, C**
Mental Component Score	43.6 (9.0)	33.3 (10.6)	22.0 (8.8)	33.23***	A***, B***, C***
HADS-A (week 1)	8.0 (2.8)	10.9 (3.1)	14.3 (3.2)	29.38***	A***, B***, C***
HADS-A (week 5)	8.5 (2.8)	10.8 (3.5)	13.7 (3.6)	13.87***	A*, B***, C**
HADS-D (week 1)	6.3 (2.9)	8.5 (3.5)	12.7 (3.1)	27.67***	A**, B***, C***
HADS-D (week 5)	5.1 (2.9)	8.7 (3.5)	9.9 (3.8)	14.34***	A***, B***

[a] HAM-A severity groups: mild = 8–14; moderate = 15–23; severe ≥ 24.
[b]Scheffe's *post hoc* pairwise comparisons: A = mild versus moderate, B = mild versus severe, C = moderate versus severe.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Table 4 Concordance between HAM-A severity groups and clinician global ratings on the CGI-S

		HAM-A severity ratings[a]			
Week	CGI-S rating	No/minimal	Mild	Moderate	Severe
Week 1 ($N = 144$)	Not at all ill	2 (100.0%)	–	–	–
	Borderline ill	4 (50.0%)	4 (50.0%)	–	–
	Mildly ill	1 (3.0%)	25 (75.8%)	7 (21.2%)	–
	Moderately ill	–	3 (3.8%)	67 (84.8%)	9 (11.4%)
	Markedly ill	–	–	6 (27.3%)	16 (72.7%)
	Severely ill	–	–	–	–
	Among the most extremely ill patients	–	–	–	–
Week 5 ($N = 109$)	Not at all ill	–	1 (100.0%)	–	–
	Borderline ill	8 (61.5%)	5 (38.5%)	–	–
	Mildly ill	3 (8.3%)	21 (58.3%)	12 (33.3%)	–
	Moderately ill	–	3 (6.3%)	36 (75.0%)	9 (18.8%)
	Markedly ill	–	–	1 (9.1%)	10 (90.9%)
	Severely ill	–	–	–	–
	Among the most extremely ill patients	–	–	–	–

[a] HAM-A scores for the four severity groups: no/minimal = 0–7; mild = 8–14; moderate = 15–23; severe ≥ 24. Statistics in this table are the frequency of patients in each cell, followed by the percentage within each row. For example, at week 1, 75.8% of patients rated as mildly ill on the CGI-S had a HAM-A score in the mild range.

be 'mildly ill' were predominantly classified in the mild HAM-A group ($n = 25$; 75.8% of patients rated by clinicians as 'mildly ill'). Patients who received clinician ratings of 'moderately ill' were predominately categorized in moderate HAM-A group ($n = 67$; 84.8% of patients rated by clinicians as 'moderately ill'). Finally, participants whom clinicians considered to be 'markedly ill' were predominately classified in the severe HAM-A group ($n = 16$; 72.7% of patients rated by clinicians as 'markedly ill'). Results followed a similar pattern at week 5.

Discussion

The study represents the first step towards developing a standard set of anxiety severity ranges for the HAM-A. Based on empirical analysis and clinical input, we propose the following cutoffs for interpreting HAM-A scores: 0–7 = no/minimal anxiety; 8–14 = mild anxiety; 15–23 = moderate anxiety; and 24 or greater = severe anxiety. In the current sample, the concordance of these score ranges with clinician ratings provides strong support for the validity of the proposed cutoffs (Table 4). Therefore, the cutoffs identified in the current study can now be applied and tested in subsequent studies.

Despite the encouraging results, the proposed score ranges should be considered only an initial suggestion. The choice of the severity cutoffs is critical because the resulting score ranges have a direct impact on the interpretation of treatment outcomes (Zimmerman *et al.*, 2004, 2005). Slight variations in cutoffs could result in over- or underestimating treatment effectiveness. The currently proposed cutoffs should be interpreted with caution because of an inherent limitation in the methodology used to identify the score ranges. This methodology, which has been used in several previous studies to identify severity ranges of other measures (Jensen *et al.*, 2001; Paul *et al.*, 2005; Serlin *et al.*, 1995; Zelman *et al.*, 2003, 2005), involves comparison among a series of models in which a three-level independent variable represents the mild, moderate, and severe groups determined by the various cutoff sets. Then, F statistics of the models are compared, with the assumption that the set of cutoffs associated with the greatest F statistic is the optimal set because it theoretically represents the greatest distinction among the mild, moderate, and severe groups. A potential problem with this assumption is that F values may be numerically different, but not meaningfully different, from each other. For example, cutoff set 1 in the current study had higher F values than cutoff set 4 (i.e. Wilk's lambda F values of 6.75 and 6.25, respectively). Thus, set 1 is considered 'better' than set 4 based on the standard methodology. However, it is not known whether the F value of 6.75 is meaningfully different from 6.25. Because of this

limitation, the current study also considered expert input (i.e. the advisory board panel) and clinician ratings (i.e. CGI-S) when determining the cutoffs, rather than relying exclusively on empirical results. Still, because it is not possible to identify score ranges with absolute certainty, the proposed cutoffs should be validated in additional samples prior to widespread acceptance and use.

The subsequent ANOVA models should also be interpreted with caution. By choosing the severity range cutoffs based on the MANOVA results, this approach inherently maximizes the chances of statistical significance in the subsequent ANOVAs, and therefore the importance of the *p* values should not be over-estimated. Still, results of these ANOVAs are useful in that they provide a more detailed explanation of the MANOVA results by presenting findings for each SF-36 and HADS subscale individually, illustrating the differences among each of the three severity groups. However, because of the circularity in this approach, these ANOVAs should not be interpreted as the only or most important validation of the cutoffs. These empirical results should be considered in combination with the clinician ratings and perceptions of the expert panel.

Additional study limitations provide directions for future research. First, the sample size of the clinical trial used in the current analysis was relatively small, particularly when dividing the sample into mild, moderate, and severe groups. Replication of this work with larger samples will add credence to the proposed score ranges. Second, when using the current method to identify cutoffs, the anchoring measures used as dependent variables in the MANOVA models have a strong impact on the results. In the current study, the combination of the HADS and SF-36 was chosen to represent anxiety symptoms and functional impact, which is similar to the approach used in previous research on other measures. For example, several previous studies used the 11-item Brief Pain Inventory, which includes a four-item pain severity scale and a seven-item scale measuring impact of pain on daily functioning (Jensen *et al.*, 2001; Serlin *et al.*, 1995; Zelman *et al.*, 2003, 2005). Future research can examine the extent to which the proposed cutoffs may vary with different dependent variables. Finally, the generalizability of the results is limited by characteristics of the sample. The current study sample consisted only of patients with GAD, and it is not known whether the resulting score ranges would be appropriate for other anxiety disorders, such as panic disorder or post-traumatic stress disorder. In addition, patients were required to meet the clinical trial inclusion/exclusion criteria, which excluded patients with common psychiatric comorbidities while limiting the range of symptom severity (as indicated by HAM-A scores) at baseline. Although there was greater variation in HAM-A scores at the time points used in the current analysis, a clinical trial sample is unlikely to be representative of the broad range of patients with GAD. Future research can build on the current study by evaluating these HAM-A severity cutoffs in broader samples.

When using the proposed cutoffs to characterize future samples, the labels for the severity groups should be applied with caution. One important purpose of these cutoffs is to facilitate communication among researchers who use the HAM-A. Therefore, we labeled the severity ranges as 'mild,' 'moderate,' and 'severe' because these terms are universally understood across clinical and research settings. However, we cannot be certain that all GAD patients categorized based on these cutoffs will be accurately labeled. In particular, the more severe range of the anxiety spectrum may not have been adequately represented in the current sample, possibly because of the relative lack of psychiatric and medical comorbidity which is typical of a clinical trial sample. Therefore, it is possible that the severe score range or label may require revision based on findings from future research conducted with a wider range of disease severity.

When using standardized patient-reported or clinician-rated instruments to assess symptom severity, the resulting scores are not meaningful unless they can be placed in context. Score cutoffs for severity ranges can provide this context, thus making results more meaningful and interpretable. If the proposed cutoffs, or similar cutoffs, are eventually accepted for general use, they will help to clarify the meaning of HAM-A scores for individual patients and their clinicians, while standardizing the interpretation of outcomes across clinical trials for GAD treatments. The current study was designed to begin this line of research, and it is hoped that future research will build on these initial results by validating the proposed cutoffs in a broad range of samples.

Acknowledgments

The authors thank the following advisors for providing input at various points during this project: Edward Schweizer, MD, John H. Griest, MD, Karl Rickels, MD, Ian Hindmarch, PhD, and Dennis A. Revicki, PhD. Julie Meilak, BA, Fritz Hamme, BA, and Aria Gray, BA, provided production assistance. This study was funded by Pfizer.

Declaration of interest statement

Louis Matza and Chris Sexton are employees of UBC, a company that received funding from Pfizer for this research. Douglas Feltner is an employee of Pfizer. Robert Morlock was an employee of Pfizer at the time this study was conducted. Karen Malley was paid by UBC for the time she spent working on this study.

References

Aben I., Verhey F., Lousberg R., Lodder J., Honig A. (2002) Validity of the Beck Depression Inventory, Hospital Anxiety and Depression Scale, SCL-90, and Hamilton Depression Rating Scale as screening instruments for depression in stroke patients. *Psychosomatics*, **43**, 386–393.

Allgulander C., Dahl A.A., Austin C., Morris P.L., Sogaard J.A., Fayyad R., Kutcher S.P., Clary C.M. (2004) Efficacy of sertraline in a 12-week trial for generalized anxiety disorder. *American Journal of Psychiatry*, **161**, 1642–1649.

Bagby R.M., Ryder A.G., Schuller D.R., Marshall M.B. (2004) The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *American Journal of Psychiatry*, **161**, 2163–2177.

Bech P., Grosby H., Husum B., Rafaelsen L. (1984) Generalized anxiety or depression measured by the Hamilton Anxiety Scale and the Melancholia Scale in patients before and after cardiac surgery. *Psychopathology*, **17**, 253–263.

Belzer K.D., Schneier F.R. (2006) Tools for assessing generalized anxiety disorder. *Psychiatric Times*, **25**(3).

Calabrese J.R., Keck P.E. Jr, Macfadden W., Minkwitz M., Ketter T.A., Weisler R.H., Cutler A.J., McCoy R., Wilson E., Mullen J. (2005) A randomized, double-blind, placebo-controlled trial of quetiapine in the treatment of bipolar I or II depression. *American Journal of Psychiatry*, **162**, 1351–1360.

Clark D.B., Donovan J.E. (1994) Reliability and validity of the Hamilton Anxiety Rating Scale in an adolescent sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, **33**, 354–360.

Davidson J.R., Bose A., Wang Q. (2005) Safety and efficacy of escitalopram in the long-term treatment of generalized anxiety disorder. *Journal of Clinical Psychiatry*, **66**, 1441–1446.

Debattista C., Solomon A., Arnow B., Kendrick E., Tilston J., Schatzberg A.F. (2005) The efficacy of divalproex sodium in the treatment of agitation associated with major depression. *Journal of Clinical Psychopharmacology*, **25**, 476–479.

Gulseren S., Gulseren L., Hekimsoy Z., Cetinay P., Ozen C., Tokatlioglu B. (2006) Depression, anxiety, health-related quality of life, and disability in patients with overt and subclinical thyroid dysfunction. *Archives of Medical Research*, **37**, 133–139.

Hamilton M. (1959) The assessment of anxiety states by rating. *British Journal of Medical Psychology*, **32**, 50–55.

Hamilton M. (1969) Diagnosis and rating of anxiety. *British Journal of Medical Psychology*, **3**(special issue), 76–79.

Jensen M.P., Smith D.G., Ehde D.M., Robinsin L.R. (2001) Pain site and the effects of amputation pain: further clarification of the meaning of mild, moderate, and severe pain. *Pain*, **91**, 317–322.

Kellner R., Kelly A.V., Sheffield B.F. (1968) The assessment of changes in anxiety in a drug trial: a comparison of methods. *British Journal of Psychiatry*, **114**, 863–869.

Lenox-Smith A.J., Reynolds A. (2003) A double-blind, randomised, placebo controlled study of venlafaxine XL in patients with generalised anxiety disorder in primary care. *British Journal of General Practice*, **53**, 772–777.

Lenze E.J., Mulsant B.H., Shear M.K., Dew M.A., Miller M.D., Pollock B.G., Houck P., Tracey B., Reynolds C.F. 3rd (2005) Efficacy and tolerability of citalopram in the treatment of late-life anxiety disorders: results from an 8-week randomized, placebo-controlled trial. *American Journal of Psychiatry*, **162**, 146–150.

Llorca P.M., Spadone C., Sol O., Danniau A., Bougerol T., Corruble E., Faruch M., Macher J.P., Sermet E., Servant D. (2002) Efficacy and safety of hydroxyzine in the treatment of generalized anxiety disorder: a 3-month double-blind study. *Journal of Clinical Psychiatry*, **63**, 1020–1027.

Maier W., Buller R., Philipp M., Heuser I. (1988) The Hamilton Anxiety Scale: reliability, validity and sensitivity to change in anxiety and depressive disorders. *Journal of Affective Disorders*, **14**, 61–68.

Marascuilo L.A., Levin J.R. (1983) *Multivariate Statistics in the Social Sciences: A Researcher's Guide*, Brooks/Cole Publishers.

Mathew S.J., Amiel J.M., Coplan J.D., Fitterling H.A., Sackeim H.A., Gorman J.M. (2005) Open-label trial of riluzole in generalized anxiety disorder. *American Journal of Psychiatry*, **162**, 2379–2381.

McHorney C.A., Ware J.E. Jr, Lu J.F., Sherbourne C.D. (1994) The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care*, **32**, 40–66.

Moller H.J. (2001) Methodological aspects in the assessment of severity of depression by the Hamilton Depression Scale. *European Archives of Psychiatry and Clinical Neuroscience, Supplement*, **251**(Suppl. 2), II13–II20.

Montgomery S.A., Sheehan D.V., Meoni P., Haudiquet V., Hackett D. (2002) Characterization of the longitudinal course of improvement in generalized anxiety disorder during long-term treatment with venlafaxine XR. *Journal of Psychiatric Research*, **36**, 209–217.

Musselman D.L., Somerset W.I., Guo Y., Manatunga A.K., Porter M., Penna S., Lewison B., Goodkin R., Lawson K., Lawson D., Evans D.L., Nemeroff C.B. (2006) A double-blind, multicenter, parallel-group study of paroxetine, desipramine, or placebo in breast cancer patients (stages I, II, III, and IV) with major depression. *Journal of Clinical Psychiatry*, **67**, 288–296.

Naukkarinen H., Raassina R., Penttinen J., Ahokas A., Jokinen R., Koponen H., Lepola U., Kanerva H., Lehtonen L., Pohjalainen T., Partanen A., Maki-Ikola O., Rouru J. (2005) Deramciclane in the treatment of generalized anxiety disorder: a placebo-controlled, double-blind, dose-finding study. *European Neuropsychopharmacology*, **15**, 617–623.

Paul S.M., Zelman D.C., Smith M., Miaskowski C. (2005) Categorizing the severity of cancer pain: further exploration of the establishment of cutpoints. *Pain*, **113**, 37–44.

Pollack M.H., Rapaport M.H., Fayyad R., Otto M.W., Nierenberg A.A., Clary C.M. (2002) Early improvement predicts endpoint remission status in sertraline and placebo treatments of panic disorder. *Journal of Psychiatric Research*, **36**, 229–236.

Rickels K., Pollack M.H., Feltner D.E., Lydiard R.B., Zimbroff D.L., Bielski R.J., Tobias K., Brock J.D., Zornberg G.L., Pande A.C. (2005) Pregabalin for treatment of generalized anxiety disorder: a 4-week, multicenter, double-blind, placebo-controlled trial of pregabalin and alprazolam. *Archives of General Psychiatry*, **62**, 1022–1030.

Rickels K., Zaninelli R., McCafferty J., Bellew K., Iyengar M., Sheehan D. (2003) Paroxetine treatment of generalized anxiety disorder: a double-blind, placebo-controlled study. *American Journal of Psychiatry*, **160**, 749–756.

Riskind J.H., Beck A.T., Brown G., Steer R.A. (1987) Taking the measure of anxiety and

depression. Validity of the reconstructed Hamilton scales. *Journal Nervous and Mental Disease*, **175**, 474–479.

Serlin R.C., Mendoza T.R., Nakamura Y., Edwards K.R., Cleeland C.S. (1995) When is cancer pain mild, moderate or severe? Grading pain severity by its interference with function. *Pain*, **61**, 277–284.

Shear M.K., Vander Bilt J., Rucci P., Endicott J., Lydiard B., Otto M.W., Pollack M.H., Chandler L., Williams J., Ali A., Frank D.M. (2001) Reliability and validity of a structured interview guide for the Hamilton Anxiety Rating Scale (SIGH-A). *Depression and Anxiety*, **13**, 166–178.

Sheehan D.V. (2001) Attaining remission in generalized anxiety disorder: venlafaxine extended release comparative data. *Journal of Clinical Psychiatry, Supplement*, **62**(19), 26–31.

Sheehan D.V., Lecrubier Y., Sheehan K.H., Amorim P., Janavs J., Weiller E., Hergueta T., Baker R., Dunbar G.C. (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, Supplement*, **59**(20), 22–33; quiz 4–57.

Shelton R.C., Keller M.B., Gelenberg A., Dunner D.L., Hirschfeld R., Thase M.E., Russell J., Lydiard R.B., Crits-Cristoph P., Gallop R., Todd L., Hellerstein D., Goodnick P., Keitner G., Stahl S.M., Halbreich U. (2001) Effectiveness of St John's wort in major depression: a randomized controlled trial. *Journal of the American Medical Association*, **285**, 1978–1986.

Smeraldi E. (1998) Amisulpride versus fluoxetine in patients with dysthymia or major depression in partial remission: a double-blind, comparative study. *Journal of Affective Disorders*, **48**, 47–56.

Wan G.J., Zhang H.F., Tedeschi M.A., Hackett D. (2006) Estimation of symptom-free days in generalized anxiety disorder. *Current Medical Research and Opinion*, **22**, 587–591.

Ware J.E. Jr, Sherbourne C.D. (1992) The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, **30**, 473–483.

Ware J.E., Snow K.K., Kosinski M.K., Gandek B. (1993) *SF-36 Health Survey: Manual and Interpretation Guide*, Health Institute, *New England Medical Center.*

Wellington K., Perry C.M. (2001) Venlafaxine extended-release: a review of its use in the management of major depression. *CNS Drugs*, **15**, 643–669.

Zelman D.C., Dukes E., Brandenburg N., Bostrom A., Gore M. (2005) Identification of cut-points for mild, moderate and severe pain due to diabetic peripheral neuropathy. *Pain*, **115**, 29–36.

Zelman D.C., Hoffman D.L., Seifeldin R., Dukes E.M. (2003) Development of a metric for a day of manageable pain control: derivation of pain severity cut-points for low back pain and osteoarthritis. *Pain*, **106**, 35–42.

Zigmond A.S., Snaith R.P. (1983) The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, **67**, 361–370.

Zimmerman M., Posternak M.A., Chelminski I. (2004) Implications of using different cut-offs on symptom severity scales to define remission from depression. *International Clinical Psychopharmacology*, **19**, 215–220.

Zimmerman M., Posternak M.A., Chelminski I. (2005) Is the cutoff to define remission on the Hamilton Rating Scale for Depression too high? *Journal of Nervous and Mental Disease*, **193**, 170–175.

Int. J. Methods Psychiatr. Res. 19(4): 223–232 (2010). DOI: 10.1002/mpr